#	LS

HKUST – Department of Computer Science and Engineering

COMP170: Discrete Math Tools for CS – FALL 2006

Final Examination

Date: Th Dec 14, 2006 Time: 12:30–3:00pm

Sketch Solution Key

Name: ____________	Student ID: ____________
Email: ____________	Lecture and Tutorial: ________

Instructions

- This is a closed book examination. It consists of 24 pages and 11 questions.
- Please write your name, student ID, Email, lecture and tutorial sections on this page.
- For each subsequent page, please write your student ID at the top of the page in the space provided.
- Please sign the honor code statement on page 2.
- Answer all the questions within the space provided on the examination paper. You may use the back of the pages for your rough work. The last three pages are scrap paper and may also be used for rough work. Each question is on a separate page (and sometimes has an extra page for you to do work on). This is for clarity and is not meant to imply that each question requires a full page answer. Many can be answered using only a few lines.
- Only use notation given in class. Do not use notation that you have learnt outside of this class that is nonstandard.
- Calculators may be used for the examination.

Questions	1	2	3	4	5	6	7	8	9	10	11	Total
Points	5	8	10	7	6	9	8	10	8	12	17	100
Score												

Student ID: ______________

As part of HKUST's introduction of an honor code, the HKUST Senate has recommended that all students be asked to sign a brief declaration printed on examination answer books that their answers are their own work, and that they are aware of the regulations relating to academic integrity. Following this, please read and sign the declaration below.

> I declare that the answers submitted for this examination are my own work.
>
> I understand that sanctions will be imposed, if I am found to have violated the University regulations governing academic integrity.
>
> Student's Name: ______________________
>
> Student's Signature: ______________________

Student ID: ______________

<u>Definitions and Formulas:</u> This page contains some definitions used in this exam and a list of formulas (theorems) that you may use in the exam (without having to provide a proof). Note that you might not need all of these formulas on this exam.

<u>Definitions:</u>

1. $N = \{0, 1, 2, 3, \ldots\}$, the set of non-negative integers.
2. $Z^+ = \{1, 2, 3, \ldots\}$, the set of positive integers.
3. R is the set of *real numbers*.
4. R^+ is the set of positive *real numbers*.

<u>Formulas:</u>

1. $\binom{n}{i} = \frac{n!}{i!\,(n-i)!}$
2. If $0 < i < n$ then $\binom{n}{i} = \binom{n-1}{i-1} + \binom{n-1}{i}$.
3. $\neg(p \wedge q)$ is equivalent to $\neg p \vee \neg q$.
4. $\neg(p \vee q)$ is equivalent to $\neg p \wedge \neg q$.
5. $\sum_{i=1}^{n-1} i = n(n-1)/2$.
6. $\sum_{i=1}^{n-1} i^2 = \frac{2n^3 - 3n^2 + n}{6}$.
7. If $r \neq 1$ then $\sum_{i=0}^{n-1} r^i = \frac{1-r^n}{1-r}$
8. If $r \neq 1$ then $\sum_{i=0}^{n} i r^i = \frac{nr^{n+2} - (n+1)r^{n+1} + r}{(1-r)^2}$
9. The inclusion-exclusion theorem:
$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{i_1, i_2, \ldots, i_k: \\ 1 \le i_1 < i_2 < \ldots < i_k \le n}} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_k})$$
10. If X is a random variable, then $E(X)$ denotes the *Expectation of* X and $V(X) = E\left((X - E(X))^2\right)$ denotes the *Variance of* X.
11. $f(n) = O(g(n))$ if there exist some $N > 0$ and positive constant c such that $\forall n > N$, $f(n) \le c \cdot g(n)$.
12. $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

Problem 1: (5pts)

For the multiple-choice problems below, **circle** the correct answer.
For this question, no work needs to be shown.

(a) How many elements in Z_{61} have a multiplicative inverse?

(i) 1 (ii) 30 (iii) 45 (iv) 60 (v) 61

(b) How many elements in Z_{77} have a multiplicative inverse?

(i) 1 (ii) 33 (iii) 60 (iv) 76 (v) 77

(c) Suppose the logical statement $\neg p \ \vee \ \neg q$ is TRUE.
Which of the following is then always also TRUE?

(i) $p \ \Rightarrow \ q$ (ii) $p \ \Rightarrow \ \neg q$ (iii) $\neg q \ \Rightarrow \ p$ (iv) $\neg q \ \Rightarrow \ \neg p$

(v) None of the above

(d) Consider the following three statements.

A) $\exists y \in N \ \Big(\ \forall x \in N \ (xy \leq x) \ \Big)$

B) $\forall x \in R \ \Big(\ \forall y \in R \ (\ (x > y) \ \Rightarrow \ (x^2 > y^2) \ ) \ \Big)$.

C) $\forall x \in R \ \Big(\ \forall y \in R \ (\ (x > y) \ \Rightarrow \ (x^3 > y^3) \ ) \ \Big)$.

Which of the three is TRUE?

(i) Only A (ii) Only B (iii) Only A and B (iv) Only A and C

(iv) All of them

(e) Suppose $0 \leq a \leq b \leq c$ and $a + b + c = n$. You are given n chairs to color. How many different ways are there to color a chairs red, b chairs blue and c chairs green?

(i) 2^{a+b} (ii) 3^n (iii) $\binom{n}{a}\binom{n-a}{b}$ (iv) $\binom{n}{a}\binom{n-b}{c}$

(v) None of the above

a (iv);
b (iii);
c (ii);
d (iv) (Only A and C);
e (iii)

Student ID: ____________

Problem 2: (8pts)

For this problem, it is not necessary to show your work.
The problem is to calculate the numbers of various hands of cards.

Recall that a card in a standard deck has two attributes:
i) 13 possible *values* or *kinds* in the order **A,2,3,4,5,6,7,8,9,10,J,Q,K**
ii) 4 suits: Diamond, Spade, Heart, Clubs
All $52 = 13 \cdot 4$ possible cards exist in the standard deck.

In this problem a *hand of 5 cards* means a subset of size 5 out of the 52 cards. So, for example, there is exactly one hand that contains all four **A**(ce)s and the **K**(ing) of Clubs.

Note the following definitions:

A *pair* is two cards of the same kind but of different suits.
Three of a kind is three cards of the same kind but of different suits.
Four of a kind is four cards of the same kind but of different suits.
A *full house* is five cards containing a *three of a kind* and another *pair* of a different value.

For example, a hand containing three **5**s and two **7**s is a full house.

Finally, Diamonds and Hearts are *red* cards while Spades and Clubs are *black*. We say that a pair is *good* if both cards in the pair are the same color, i.e., they are both red or they are both black.

1. How many hands of 5 cards contain a *four of a kind*?
2. How many hands of 5 cards contain two *good pairs*? (*Four of a kind* is considered as two good pairs.)
3. How many hands of 5 cards contain two *pairs* but do not contain *four of a kind* and do not contain a *full house*?

1. *You can first choose the value of the four of the kind and then choose the other card. This give* $13 \cdot 48 = 624$
2. *There are 26 possible good pairs so there are* $\binom{26}{2}$ *possible ways to choose the two good pairs. Choosing the final card gives* $\binom{26}{2} \cdot 48 = 15,600$.
3. *There are* $13\binom{4}{2}$ *ways of choosing the "first" pair and* $12\binom{4}{2}$ *ways of choosing the "second" pair. Since order doesn't count here you have to divide by 2. Then multiply by the* 44 *ways of choosing the non-pair card to get*
$$\frac{13\binom{4}{2}12\binom{4}{2}}{2}44 = 123,552$$

Student ID: ____________

Problem 3: (10pts)

In this problem you do not *have* to show your work. You may assume that a, b, c are all non-negative integers.

The running time of an algorithm A is described by the following recurrence relation:

$$S(n) = \begin{cases} b & n = 1 \\ 9S(n/2) + n^2 & n > 1 \end{cases}$$

where b is a positive constant and n is a power of 2. The running time of a competing algorithm B is described by the following recurrence relation:

$$T(n) = \begin{cases} c & n = 1 \\ aT(n/4) + n^2 & n > 1 \end{cases}$$

where a and c are two positive constants and n is a power of 4.

For the rest of this problem you may assume that n is always a power of 4. You should also assume that $a > 16$.

1. Find a solution for $S(n)$. Your solution should be in closed form (in terms of b if necessary) and should *not* use summation.
2. Find a solution for $T(n)$. Your solution should be in closed form (in terms of a and c if necessary) and should *not* use summation.
3. Algorithm B is said to be *at least as efficient as Algorithm A asymptotically* (i.e., as n becomes very large) if $T(n) = O(S(n))$. For what range of values of $a > 16$ is Algorithm B at least as efficient as Algorithm A asymptotically?

1. *By repeated substitution, we get*

$$\begin{aligned}
S(n) &= 9S\left(\frac{n}{2}\right) + n^2 \\
&= 9\left[9S\left(\frac{n}{2^2}\right) + \left(\frac{n}{2}\right)^2\right] + n^2 \\
&= 9^2 S\left(\frac{n}{2^2}\right) + \left(\frac{9}{4}\right) n^2 + n^2 \\
&= 9^2\left[9S\left(\frac{n}{2^3}\right) + \left(\frac{n}{2^2}\right)^2\right] + \left(\frac{9}{4}\right) n^2 + n^2 \\
&= 9^3 S\left(\frac{n}{2^3}\right) + \left(\frac{9}{4}\right)^2 n^2 + \left(\frac{9}{4}\right) n^2 + n^2 \\
&= \cdots \\
&= 9^{\log_2 n} S(1) + n^2 \sum_{i=0}^{\log_2 n - 1} \left(\frac{9}{4}\right)^i \\
&= b\, n^{\log_2 9} + \frac{4}{5} n^{\log_2 9} - \frac{4}{5} n^2 \\
&= \left(b + \frac{4}{5}\right) n^{\log_2 9} - \frac{4}{5} n^2,
\end{aligned}$$

where we are using the fact that

$$\left(\frac{9}{4}\right)^{\log_2 n} = \frac{9^{\log_2 n}}{n^2} = \frac{n^{\log_2 9}}{n^2}.$$

2. *Similar to $S(n)$, by repeated substitution, we get*

$$\begin{aligned}
T(n) &= aT\left(\frac{n}{4}\right) + n^2 \\
&= a\left[aT\left(\frac{n}{4^2}\right) + \left(\frac{n}{4}\right)^2\right] + n^2 \\
&= a^2T\left(\frac{n}{4^2}\right) + \left(\frac{a}{16}\right)n^2 + n^2 \\
&= a^2\left[aT\left(\frac{n}{4^3}\right) + \left(\frac{n}{4^2}\right)^2\right] + \left(\frac{a}{16}\right)n^2 + n^2 \\
&= a^3T\left(\frac{n}{4^3}\right) + \left(\frac{a}{16}\right)^2 n^2 + \left(\frac{a}{16}\right)n^2 + n^2 \\
&= \cdots \\
&= a^{\log_4 n}T(1) + n^2 \sum_{i=0}^{\log_4 n-1} \left(\frac{a}{16}\right)^i \\
&= c\,n^{\log_4 a} + \frac{16}{a-16}n^{\log_4 a} - \frac{16}{a-16}n^2 \\
&= \left(c + \frac{16}{a-16}\right)n^{\log_4 a} - \frac{16}{a-16}n^2,
\end{aligned}$$

where we are using the fact that

$$\left(\frac{a}{16}\right)^{\log_4 n} = \frac{a^{\log_4 n}}{n^2} = \frac{n^{\log_4 a}}{n^2}.$$

3. *For $T(n) = O(S(n))$, we should have*

$$\begin{aligned}
n^{\log_4 a} &\leq n^{\log_2 9} \\
\log_4 a &\leq \log_2 9 \\
a &\leq 9^2 = 81.
\end{aligned}$$

So the range of values is $16 < a \leq 81$.

Student ID: ____________

Problem 4: (7pts)

Let $E_1, E_2, \ldots, E_n$ be a collection of any $n \geq 2$ events. Prove by induction on n that

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) \leq \sum_{i=1}^{n} P(E_i).$$

If you use any result that we proved in class, state the result formally.

We first prove that the inequality holds for the base case ($n = 2$). Applying the inclusion-exclusion principle to events E_1 and E_2, we have

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2).$$

Since $P(E_1 \cap E_2) \geq 0$, so

$$P(E_1 \cup E_2) \leq \sum_{i=1}^{2} P(E_i).$$

Hence the base case is true.

Next we consider the inductive step for $n > 2$. Assuming that the inequality holds for $n - 1$, i.e.,

$$P(E_1 \cup E_2 \cup \cdots \cup E_{n-1}) \leq \sum_{i=1}^{n-1} P(E_i),$$

we want to show that it also holds for n.

$$\begin{aligned}
P(E_1 \cup E_2 \cup \cdots \cup E_n) &= P((E_1 \cup E_2 \cup \cdots \cup E_{n-1}) \cup E_n) \\
&\leq P(E_1 \cup E_2 \cup \cdots \cup E_{n-1}) + P(E_n) \quad \textit{(base case)} \\
&\leq \sum_{i=1}^{n-1} P(E_i) + P(E_n) \quad \textit{(inductive hypothesis)} \\
&= \sum_{i=1}^{n} P(E_i).
\end{aligned}$$

From the principle of mathematical induction, we can conclude that the inequality above holds for all $n \geq 2$.

`Note: Students made two major mistakes when solving this problem. The first was to use an induction basis of` $n = 1$ `instead of` $n = 2$`. Although` $n = 1$`, i.e.,` $E(1) \leq E(1)$ `is true, it doesn't help in proving that the size` $n - 1$ `case implies the size` n `case.`

`The second mistake was to try to prove the statement directly from the general inclusion-exclusion formula on` n `events. This would be very difficult to do correctly.`

Problem 5: (6pts)

In this problem you only need to give the answer. It is not necessary to show your work.

Suppose a student knows 80% of the material in a course and she takes a test with 100 multiple-choice questions, each with four choices. We assume that the test covers the course material uniformly (i.e., all the questions are independent of each other) so she has an 80% chance of knowing the answer to each question.

The student is not careful, though, and even if she *knows* the answer to a question, she has a 5% chance of making a careless mistake and hence marking a wrong answer. If she does not know the material, she makes a random guess to pick one choice. Let X be her number of correct answers.

1. What is $E(X)$?
2. What is $V(X)$?

Let K_i denote the event that the student knows the material asked in the ith question and C_i the event that she gets the correct answer to the question. Thus we have the following conditional probabilities:

$$\begin{aligned} P(C_i \mid K_i) &= 0.95 \\ P(C_i \mid \overline{K}_i) &= 0.25. \end{aligned}$$

So

$$\begin{aligned} P(C_i) &= P(C_i \mid K_i)P(K_i) + P(C_i \mid \overline{K}_i)P(\overline{K}_i) \\ &= 0.95 \cdot 0.8 + 0.25 \cdot 0.2 \\ &= 0.81 \end{aligned}$$

Set $p = 0.81$. Let X_i be the indicator random variable for C_i, i.e., $X_i = 1$ if she gets question i correct and zero, otherwise, Then $E(X_i) = p$ and $V(X_i) = p(1-p)$. Finally, notice that $X = \sum_{i=1}^{100} X_i$.

1. $E(X) = \sum_{i=1}^{100} E(X_i) = 100p = 81.$
2. *Because the X_i are independent trials we find that* $V(X) = \sum_{i=1}^{100} V(X_i) = 100p(1-p) = 15.39$

Student ID: ____________

Problem 6: (9pts)

For this problem you may assume that n is a nonnegative power of 4. Suppose function $T(n)$ satisfies $T(1) = 6$ and, for $n > 1$,

$$T(n) \le 4T\left(\frac{n}{4}\right) + 2n$$

Prove that $T(n) = O(n \log_4 n)$. That is, you should prove that there exist some $n_0 \ge 0$ and $k > 0$ such that

$$\forall n > n_0, \quad T(n) \le kn \log_4 n.$$

Hint: Prove by induction.

We start by noticing that it is impossible for $n_0 = 0$ since that would mean that

$$6 = T(1) \le k \log_4 1 = 0.$$

We therefore choose $n_0 = 1$. The base of our induction will then be

$$T(4) \le 4k \log_4 4 = 4k.$$

We don't know what $T(4)$ is but, from the defining inequality, we do know that $T(4) \le 4T(1) + 2 \cdot 4 = 32$. This if we assume that (i) $k \ge 8$ we will have the base case

$$T(n) \le 4T\left(\frac{n}{4}\right) + 2n$$

with $n = 4$.

Now, suppose that the statement is correct for all $n = 4^j$, $j = 1, 2, 3, \dots i-1$. When $n = 4^i$ we use the induction hypothesis to get

$$\begin{aligned} T(n) &\le 4T\left(\frac{n}{4}\right) + 2n \\ &\le 4k\frac{n}{4}\log_4 \frac{n}{4} + 2n \\ &= kn(\log_4 n - 1) + 2n \\ &= kn \log_4 n - (k-2)n \end{aligned}$$

So as long as (ii) $k \ge 2$ we will have proven that $T(n) \le kn \log_4 n$ for $n = 4^i$. Thus, from (i) and (ii), as long as $k \ge \max\{8, 2\} = 8$ we will have $T(n) \le kn \log_2 n$ for all $n = 4^j$ where $n > 1$.

`Note: The major mistake made on this question was to choose` $n_0 =$ 0`. As discussed above, this is wrong because` $T(1) > k \log_4 1 = 0$ `for every value of` k.

Problem 7: (8pts)

Recall that a card in a standard deck has two attributes:
i) 13 possible *values* or *kinds* in the order **A,2,3,4,5,6,7,8,9,10,J,Q,K**
ii) 4 suits: Diamond, Spade, Heart, Clubs
All $52 = 13 \cdot 4$ possible cards exist in the standard deck.

Suppose that every morning you randomly pick one card from your standard deck, look at it and then return it to the deck.

Let X be the number of days until you have seen at least one card for each of the 13 different values.

What is $E(X)$? (Prove your answer).

This is very similar to the coupon collectors problem in the extra handouts. Keep a chart of the 13 values. Tick off a value the first time you pick it. Let X_i be the number of picks needed until i values have been picked. Note that $X = X_{13}$.

Set $Y_1 = 1$. For $i > 1$ set $Y_i = X_i - X_{i-1}$ be the number of picks needed from the time you tick off the $(i-1)$st value to the time you tick off the i^{th} one.

The important observations are that (i) $X = \sum_{i=1}^{13} Y_i$ and (ii) Y_i is an independent trials process with success probability $p_i = \frac{13-(i-1)}{13}$ that stops at the first success. So, $E(Y_i) = 1/p_i$ and, by linearity of expectation

$$\begin{aligned}
E(X) &= \sum_{i=1}^{13} E(Y_i) \\
&= \sum_{i=1}^{13} (1/p_i) \\
&= \sum_{i=1}^{13} \frac{13}{13-(i-1)} \\
&= 13 \sum_{i=1}^{13} \frac{1}{i}.
\end{aligned}$$

This can also be written as $E(X) = 13H_{13}$ where $H_n = \sum_{i=1}^{n} \frac{1}{i}$ is the n^{th} Harmonic number.

Problem 8: (10pts)

A new software engineer has just implemented the RSA algorithm for public key cryptography. His program generates two positive integers $p = 13$ and $q = 18$ that are relatively prime, i.e., $\gcd(p, q) = 1$. It then computes $n = pq = 234$ and $T = (p-1)(q-1) = 204$, and chooses an integer $e = 5$ that is relatively prime to T, i.e., $\gcd(e, T) = 1$.

His program then finds d, the multiplicative inverse of e in Z_T. i.e., $d \in Z_T$ such that $d \cdot e \bmod T = 1$.
The pair (n, e) is then assigned as the public key for Bob who keeps (n, d) as his corresponding private key.

1. What is the value of d?
   It is not necessary to show how you found d.
2. Alice wants to send a message $M = 3$ to Bob by applying RSA encryption. The encrypted message $\bar{M}$ is
   $$M^e \bmod n = 3^5 \bmod 234 = 9.$$
   Upon receiving the encrypted message, Bob applies RSA decryption to get
   $$\hat{M} = \bar{M}^d \bmod n = 9^d \bmod 234.$$
   Calculate $\hat{M}$. Show all of your work.
3. Unfortunately (your calculations should show) $\hat{M}$ is not equal to M! What went wrong? Doesn't RSA ensure that $\hat{M} = M$? Explain why RSA algorithm did not work in this situation.

1. $d = 41$

2. *Use repeated squaring to find*

$$\begin{aligned} 9^2 \bmod 234 &= 81 \\ 9^4 \bmod 234 &= 9 \\ 9^8 \bmod 234 &= 81 \\ 9^{16} \bmod 234 &= 9 \\ 9^{32} \bmod 234 &= 81 \end{aligned}$$

so

$$\hat{M} = 9^{41} \bmod 234 = \left(9^{32} \cdot 9^8 \cdot 9\right) \bmod 234 = (81 \cdot 81 \cdot 9) \bmod 234 = 81$$

3. RSA requires that p and q are both prime, not just relatively prime.

Problem 9: (8pts)

A modified Towers of Hanoi puzzle.

As in the standard Towers of Hanoi puzzle, there are 3 posts and n disks of different sizes. Initially, all n disks are on post 1. The objective is to transfer all n disks to post 2 via a sequence of moves. A move consists of removing the top disk from one post and dropping it onto another post, with the restriction that a larger disk can never lie above a smaller disk. Furthermore, *a disk can be moved only from post 1 to post 2, from post 2 to post 3, or from post 3 to post 1.* Thus, for example, moving a disk directly from post 1 to post 3 is not allowed.

Here is one (not necessarily optimal) solution to the problem:

- If there is only one disk, then, like the standard Towers of Hanoi problem, we only need one move to move the disk from post 1 to post 2.
- If there are two disks, we first move the smaller disk from post 1 to post 2 and then from post 2 to post 3. Afterwards, we move the larger disk from post 1 to post 2 directly. Finally, we move the smaller disk from post 3 to post 1 and then from post 1 to post 2.

This can be generalized easily to any $n > 1$ disks as follows.

- First, we (recursively) move the top stack of $n-1$ disks from post 1 to post 2 and then from post 2 to post 3.
- Afterwards, we move the largest disk from post 1 to post 2 directly.
- Finally, we (recursively) move the $n-1$ disks on post 3 to post 1 and then from post 1 to post 2.

1. Let $T(n)$ be the number of moves needed to solve the n-disk problem based on the general solution described above. Express $T(n)$ as a recurrence relation.
2. Find a closed-form expression for $T(n)$ by solving the recurrence. Your expression should *not* use summation.

1.

$$T(n) = \begin{cases} 1 & n = 1 \\ 4T(n-1) + 1 & n > 1 \end{cases}$$

2. $T(n) = (4^n - 1)/3$

Student ID: ____________

Problem 10: (12pts)

In the board game *Monopoly*, the number of squares that a player advances in a single *turn* is determined by up to three rolls of a pair of dice as follows:

- Initially, the player rolls two dice, sums the results, and advances that many squares.
- If the same number comes up on both dice—this is called *rolling a double*—then the player rolls the dice a second time, sums the results, and advances that many additional squares.
- If the player rolled a double on the second roll, then the player rolls the dice a third time, sums the results, and advances that many additional squares.
- However, if the third roll was a double, then the player "goes to jail". We will treat this as though he had to reset his position to where he was just before the first roll, so the total number of squares advanced by the player in this case is zero.

Assume that the dice are fair and six-sided and the rolls are independent trials. Answer the following four questions. In all parts you must show all the steps in getting the answer.

1. What is the probability of going to jail on a turn?
2. What is the probability that you advance 6 squares in one turn?
3. What is the probability that you advance 7 squares in one turn?
4. What is the expected number of squares that you advance in one turn?

1. *Let I_i be the indicator random variable for the event that the ith roll is a double.*
$$P(I_i = 1) = E(I_i) = \frac{6}{36} = \frac{1}{6}.$$
The probability of going to jail is equal to
$$P(I_1 I_2 I_3 = 1) = E(I_1 I_2 I_3) = E(I_1) \cdot E(I_2) \cdot E(I_3) = \frac{1}{216}.$$

2. *Let R_i be the random variable representing the sum of the numbers on the two dice at the ith roll and (r_i, s_i) be the corresponding ordered pair of numbers, i.e., $R_i = r_i + s_i$.*

   To advance 6 squares in a single turn, the following table shows all possible cases and the corresponding probabilities.

r_1	s_1	r_2	s_2	r_3	s_3	*Probability*
1	5					1/36
2	4					1/36
4	2					1/36
5	1					1/36
1	1	1	3			$(1/36)^2$
1	1	3	1			$(1/36)^2$

   The total probability is equal to $(1/36) \cdot 4 + (1/36)^2 \cdot 2 = 73/648$.

3. *To advance 7 squares in a single turn, the following table shows all possible cases and the corresponding probabilities.*

r_1	s_1	r_2	s_2	r_3	s_3	*Probability*
1	6					1/36
2	5					1/36
3	4					1/36
4	3					1/36
5	2					1/36
6	1					1/36
1	1	1	4			$(1/36)^2$
1	1	2	3			$(1/36)^2$
1	1	3	2			$(1/36)^2$
1	1	4	1			$(1/36)^2$
2	2	1	2			$(1/36)^2$
2	2	2	1			$(1/36)^2$
1	1	1	1	1	2	$(1/36)^3$
1	1	1	1	2	1	$(1/36)^3$

   The total probability is equal to $(1/36) \cdot 6 + (1/36)^2 \cdot 6 + (1/36)^3 \cdot 2 = 3997/23328$.

4. *Let R be the random variable for the number of squares advanced by the player. We can express R in terms of R_i and I_i as*

$$R = R_1 + I_1 R_2 + I_1 I_2 R_3 - I_1 I_2 I_3 (R_1 + R_2 + R_3).$$

Taking expectation on both sides, with the assumption that the rolls are independent, we have

$$\begin{aligned}
E(R) &= E(R_1 + I_1 R_2 + I_1 I_2 R_3 - I_1 I_2 I_3 (R_1 + R_2 + R_3)) \\
&= E(R_1 + I_1 R_2 + I_1 I_2 R_3 - I_1 I_2 I_3 R_1 - I_1 I_2 I_3 R_2 - I_1 I_2 I_3 R_3) \\
&= E(R_1) + E(I_1 R_2) + E(I_1 I_2 R_3) - E(I_1 I_2 I_3 R_1) - E(I_1 I_2 I_3 R_2) \\
&\quad - E(I_1 I_2 I_3 R_3) \\
&= E(R_1) + E(I_1)E(R_2) + E(I_1)E(I_2)E(R_3) - E(I_1 R_1)E(I_2)E(I_3) \\
&\quad - E(I_1)E(I_2 R_2)E(I_3) - E(I_1)E(I_2)E(I_3 R_3).
\end{aligned}$$

The expected value $E(R_i)$ can be computed as

$$E(R_i) = 2 \cdot \frac{\sum_{i=1}^{6} i}{6} = 7.$$

Also, from above, we have $E(I_i) = 1/6$. For $E(I_i R_i)$, we can compute it directly as

$$E(I_i R_i) = \frac{1}{36} \sum_{i=1}^{6} (2i) = \frac{7}{6}.$$

Substituting these into $E(R)$ above, we obtain

$$E(R) = \frac{301}{36} - \frac{7}{72} = \frac{595}{72}.$$

Problem 11: (17pts)

Throwing balls into boxes.
When we say *throw m balls into n boxes* we mean that each ball has probability $1/n$ of being thrown into each box and the different balls are independent of each other. Answer the following four questions. In all parts you must show all the steps in getting the answer.

1. Throw 6 balls into 5 boxes.
   What is the probability that every box contains at least one ball?
   Express your answer as a decimal number (with accuracy at least 4 places after the decimal point).
2. Throw 150 balls into 100 boxes. What is the probability that every box contains at least one ball?
   You should write the solution to this problem as a summation of numbers.
3. Throw 20 balls into 10 boxes.
   Let X be the number of boxes containing exactly 1 ball.
   Let Y be the number of boxes containing exactly 2 balls.
   What is $E(X+Y)$?.
   Write a formula for your solution (it should *not* use summation).
4. Throw m balls into n boxes.
   Let X_i be the number of balls in box i.
   Let $Y_i = \sum_{j=1}^{i} X_j$ be the number of balls in the first i boxes.
   Write a formula expressing $V(Y_i)$ in terms of m, n, and i.
   Your solution should not use summation.

1. *The number of ways to throw 6 balls into 5 boxes so that every box contains one ball is exactly the same as the number of onto functions from* $\{1, 2, 3, 4, 5, 6\}$ *to* $\{1, 2, 3, 4, 5\}$ *which is* $\binom{6}{2}5!$. *There are* 5^6 *different ways of throwing the 6 balls into 5 boxes so the probability is*

$$\frac{\binom{6}{2}5!}{5^6}.$$

2. *Let* A *be the event that every box contains at least 1 ball. Let* B *be the event that at least 1 box is empty. Since* B *is the complement of* A, *we have* $P(A) = 1 - P(B)$.

   Now, for $i = 1, \ldots 100$ *set* E_i *to be the event that box* i *is empty. Note that* $B = \bigcup_{i=1}^{100} E_i$ *so we can use the inclusion-exclusion formula*

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=1}^{n}(-1)^{k+1} \sum_{\substack{i_1, i_2, \ldots, i_k: \\ 1 \le i_1 < i_2 < \ldots < i_k \le n}} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_k})$$

   to calculate $P(B)$.

   The first thing to note is that, for fixed k,

$$P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_k}) = \left(1 - \frac{k}{100}\right)^{150}$$

   There are $\binom{100}{k}$ *ways of choosing such* k*-tuples so*

$$\sum_{\substack{i_1, i_2, \ldots, i_k: \\ 1 \le i_1 < i_2 < \ldots < i_k \le n}} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_k}) = \binom{100}{k}\left(1 - \frac{k}{100}\right)^{150}$$

   Thus

$$P(B) = \sum_{i=1}^{100}(-1)^{k+1}\binom{100}{k}\left(1 - \frac{k}{100}\right)^{150}.$$

   Note that when $k = 100$ *the term is equal to zero so we can run the summation from 1 to 99 instead of 1 to 100 if we so wanted.*

   Now, $P(A) = 1 - P(B)$ *and we are done.*

3. *Let* X_i *be the indicator random variable for box* i *containing exactly 1 ball and* Y_i *the indicator random variable for box* i *containing exactly two balls. Then*

$$X = \sum_{i=1}^{10} X_i, \quad Y = \sum_{i=1}^{10} Y_i.$$

   Now note that

$$E(X_i) = 20\frac{1}{10}\left(\frac{9}{10}\right)^{19}, \quad E(Y_i) = \binom{20}{2}\left(\frac{1}{10}\right)^{2}\left(\frac{9}{10}\right)^{18}.$$

Then

$$E(X+Y) = \sum_{1=1}^{10} E(X_i) + \sum_{1=1}^{10} E(Y_i) = 10 \cdot 20 \frac{1}{10} \left(\frac{9}{10}\right)^{19} + 10 \cdot \binom{20}{2} \left(\frac{1}{10}\right)^{2} \left(\frac{9}{10}\right)^{18}$$

4. *Let Z_j be the event that ball j falls into one of the first i boxes. Then $E(Z_j) = \frac{i}{n}$ and $V(Z_j) = \frac{i}{n}\left(1 - \frac{i}{n}\right)$.*
The next thing to notice is that $Y_i = \sum_{j=1}^{m} Z_j$ and all of the Z_j are independent so

$$V(Y_i) = \sum_{j=1}^{m} V(Z_j) = m\frac{i}{n}\left(1 - \frac{i}{n}\right).$$

Note: Some students wrote that since, for all j, X_j is binomial with $p = \frac{1}{n}$ we have $V(X_j) = \frac{m}{n}\left(1 - \frac{1}{n}\right)$. We can therefore write

$$V(Y_i) = \sum_{j=1}^{i} V(X_j) = iV(X_1) = i\frac{m}{n}\left(1 - \frac{1}{n}\right).$$

This is NOT *correct. The reason is that the X_i are not independent so we do not necessarily have $V(Y_i) = \sum_{j=1}^{i} V(X_j)$.*